THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF APPLIED BIOLOGY &
CHEMICAL TECHNOLOGY

Kernel Based Approaches in Drug-Target Interaction Prediction

Xinyi ZENG

Submitted for the degree of Bachelor of Science

The Hong Kong Polytechnic University

April 2016

Contents

List of Tables

Abstract

In silico prediction of drug-target interactions effectively facilitate drug candidates selecting process prior to biochemical verification. Machine learning had its heyday in drug discovery around two decades ago. To date, we are witnessing renewed interest in kernel learning, a branch of machine learning that enjoys outstanding performance by integrating chemical and genomic information for prediction. This article provide an overview of kernel based approaches in drug-target interaction prediction, present each of these frameworks with distinctive kernels and subsequently analyse their strengths as well as restrictions.

Keywords: drug-target interaction prediction; machine learning; drug discovery; kernel method

Chapter 1

Introduction

The identification of interactions between compounds and therapeutic target proteins typified by enzymes, G-protein-coupled receptors (GPCR), ion channels and nuclear receptors is essential in genomic drug discovery [?]. Heterogeneous genomic spaces revealed with the completion of human genome sequencing project and numerical biological assays cached in chemical compound databases allowing high-throughput screening bring strong incentive to researches on correlating genomic space to chemical space for lead discovery.

With various latent interactions still remaining murky, in silico approach is increasingly gaining research attention for its prediction power and auxiliary role in experimental studies by providing optimisation route and supporting evidences. Classical computational methods including text mining and molecular docking fail to concurrently exploit information of chemical space and protein sequence besides their own inevitable restrictions, molecular docking, for example, is highly dependent on 3D structures of proteins. Text mining loses it efficacy on latest discoveries and redundancy of nomenclature [?]. Kernel learning algorithms are developed progressively for its remarkable performance achieved by taking both aspects into consideration.

A kernel can be considered as a similarity function which outputs the analogy between compounds or proteins. Kernel methods construct a rational framework to integrate data sources in various types which enhances data fusion capability of

statistical learning methods in order to model a joint relationship between chemical space, genomic sequence along with drug-target interaction pattern and thus accuracy improves.

The objective of this literature review is two-fold: 1) to critically summarise the evolution of kernel methods in recent years and therefore 2) to motivate further development of kernel based models.

Chapter 2

Problem formulation

For a given set of N_c drug compounds $\mathbf{D} = \{d_1, d_2, \ldots, d_{N_c}\}$, another set of N_t target proteins $\mathbf{T} = \{t_1, t_2, \ldots, t_{N_t}\}$ and a $N_c \times N_t$ binary matrix $\mathbf{Y}$ whose (i,j)-th entry is labelled as +1 if drug d_i targets on protein t_j and labelled as -1/0 otherwise which will be further discussed under specific condition, the prediction of drug-target interaction problem can be categorised into three scenarios:

a. to find targeted proteins in $\mathbf{T}$ for an out-of-sample drug $d^{'}$,

b. to find compounds in $\mathbf{D}$ that target a new protein $t^{'}$,

c. to theorising an unknown interaction of the network between drug $d^{'}$ and target $t^{'}$.

2.1 Support vector machine

It is intuitively to address the interaction prediction issue using binary classification methods where class +1 refers to interaction and class -1 to non-interaction - nearest neighbour classifiers, decision trees, and support vector machines (SVM), to name a few. SVM is regarded highly among chemoinformatists and bioinformatists for its capability to handle structured data for classification.

A kernalised SV classifier typically learn a hyperplane by

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b \tag{2.1}$$

from N pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ with $y_i \in \{-1, +1\}$, k is a kernel function outputs the similarity between two inputs. For separable data, there exists a hyperplane with a maximal margin M between two classes. Boundaries can be softened by introducing a slack variable b if data is inseparable. The design of kernel functions borrows information from chemical structures of molecules and sequence information of proteins. With the only constraint to be a symmetrically positive definite function, various kernel functions can be employed [**?**].

Pairwise kernels

Vectors representations for drug d $\Phi_{lig}(c) \in \mathbb{R}^{d_c}$, target t $\Phi_{tar}(t) \in \mathbb{R}^{d_t}$ and each pair (t, d) $\Phi(d, t)$ informative enough to capture the physico-chemical and structural properties of molecules and the sequence features of proteins are crucial in order to utilise statistical methods to predict interactions between numerous targets for any given molecule. The vector representation of a drug-target pair $\Phi(d, t)$ can be a direct addition of $\Phi_{lig}(c)$ and $\Phi_{tar}(t)$ [**?**]:

$$\Phi(d, t) = (\Phi_{lig}(d)^T, \Phi_{tar}(t)^T)^T \tag{2.2}$$

or the tensor product [**?**]:

$$\Phi(d, t) = \Phi_{lig}(d) \otimes \Phi_{tar}(t) \tag{2.3}$$

where the (i,j)-th element is the product of the i-th element of $\Phi_{lig}(d)$ and the j-th element of $\Phi_{tar}(t)$.

From the perspective of computational burden, **?** incorporate feature vectors by simple addition resulting a size of $(d_c + d_t)$ rather than a size of $(d_c \times d_t)$ for tensor product. However, the second approach leads to more flexible combinations of vector representations of small compounds and proteins. It is well worth congratulating that **?** significantly reduces the computational complexity by factorising the inner product between two tensor product vectors:

$$(\Phi_{lig}(d) \otimes \Phi_{tar}(t))^T (\Phi_{lig}(d') \otimes \Phi_{tar}(t')) = \Phi_{lig}(d)^T \Phi_{lig}(c') \times \Phi_{tar}(t)^T \Phi_{tar}(t') \tag{2.4}$$

so that only the two inner products in dimensions d_c and d_t need computing before taking their product. With two kernels:

$$K_{lig}(d, d') = \Phi_{lig}(d)^T \Phi_{lig}(d') \qquad (2.5)$$

$$K_{tar}(t, t') = \Phi_{tar}(t)^T \Phi_{tar}(t') \qquad (2.6)$$

that map chemical structures and genomic sequences into two Hilbert spaces, the inner product between two tensor products can be derived:

$$K((d, t), (d', t')) = K_{lig}(d, d') \times K_{tar}(t, t') \qquad (2.7)$$

which serves as an input in an SVM. With only one constraint to be a positive semi-definite function for a kernel function, a graph kernel is considered to be a capable candidate for a small molecule that encodes its relevant properties and string kernels for proteins.

Bipartite local model

As an extension of local model approaches [**??**], two local models are trained to give independent predictions for the same edge whose existence indicates an interaction between a drug node d_i and a target node t_j on a bipartite graph from two subsets with either drug d_i or target t_j excluded from an original dataset[**?**]. In the subset excluding drug d_i, +1 is assigned to the drugs targeting t_j or -1 otherwise so that a classifier discriminating two types of drugs predicts the existence of the edge on a scale of -1 to +1. Comparing with another score obtained from the subset without target t_j, the higher value is assigned to be the final score to lighten the effect of -1 data due to a major defect in bipartite local model (BLM) since uncertain edges have been assumed non-interaction previously and thus cannot be deduced. SVMs are chosen as local models for the advantage of providing continuous scores and handy kernel tricks.

2.2   Regression method

Predictions of a bipartite graph learning method [**?**] is based on the similarity of features calculated from three inner products: correlations of 1) the new molecule

with the target, 2) a studied molecule with the new target and 3) the new molecule with the new target by considering the chemical reaction as a pattern. It consists of three procedures:

1. Similarity matrix

   A graph-based similarity matrix $M = \left(\begin{smallmatrix} M_{dd} & M_{dt} \\ M_{dt}^{T} & M_{tt} \end{smallmatrix} \right)$, whose entries are calculated using Gaussian functions:

   $$(M_{dd})_{ij} = exp(\frac{-d_{d_i d_j}^2}{h^2}), i, j = 1, \ldots, n_d \tag{2.8}$$

   $$(M_{dt})_{ij} = exp(\frac{-d_{d_i t_j}^2}{h^2}), i = 1, \ldots, n_d, j = 1, \ldots, n_t \tag{2.9}$$

   $$(M_{tt})_{ij} = exp(\frac{-d_{t_i t_j}^2}{h^2}), i, j = 1, \ldots, n_t \tag{2.10}$$

   where:

        d: The shortest distance between molecules or proteins on a bipartite graph

        h: The width parameter

   with an identity matrix K_{dt}^T to guarantee the positive definite property in order to represent the drug-target interaction in an Euclidian Space using feature vectors $\{u_{d_i}\}_{i=1}^{n_d}$ and $\{u_{t_i}\}_{j=1}^{n_t}$. Analogous with the component analysis proposed by **?**, eigenvalue decomposition is applied so that all the drugs and targets can be represented by row vectors of matrix $U = (u_{d_1,\ldots,u_{d_{n_d}}}, u_{t_1}, \ldots, u_{t_{n_t}})^T$

2. Regression model

   Two kernelised regression models that correlate the chemical spaces of molecules or genomic sequences of targets with the pharmacological feature space to project the coming molecules or targets into the Euclidean space:

   $$u_{d'} = f_d(d', d_i) \quad = \sum_{i=1}^{n_d} S_d(d', d_i) w_{d_i} \tag{2.11}$$

   $$u_{t'} = f_t(t', t_i) \quad = \sum_{i=1}^{n_t} S_t(t', t_i) w_{t_i} \tag{2.12}$$

   where:

        S: the similarity matrix

w: a weight vector

are learned from minimising the loss function:

$$L = ||UU^T - SWW^TS^T||_F^2 \qquad (2.13)$$

where:

S: a similarity matrix

$W = (w_1, \ldots, w_n)^T$

$|| \cdot ||_F$: Frobenius norm

3. Prediction

   Drug-target pairs with closeness indicated by the inner product between corresponding molecules and targets in the pharmacological feature space larger than a threshold ϵ are labelled as non-interaction and vice versa.

2.3 Regularised least square

Network-based Laplacian RLS

To overcome the scarcity of known drug-target interactions in biochemical data, semi-suppervised learning methods (SSL) not only discover knowledge from myriad unlabelled data besides labelled one but also combine interaction patterns with chemical structures and genomic sequence information in order to achieve more accurate predictions. **?** implemented this framework using Laplacian regularised least squares (LapRLS), a basic classifier with comparable accuracy with SVM if proper kernels are employed, to train two classifiers for a drug d' with a target t_j and a target t' with a drug d_i separately

$$f(d', t_j) = \frac{1}{N_d} \sum_{i=1}^{n_d} s_d(d', d_i) \mathbf{Y}_{ij} \qquad (2.14)$$

$$f(t', d_i) = \frac{1}{N_t} \sum_{j=1}^{n_t} s_d(t', t_j) \mathbf{Y}_{ij} \qquad (2.15)$$

where:

$\mathbf{Y}$: a binary matrix with assured interactions between drug d_i and target t_j labelled as 1 and undiscovered pairs as 0 to be further predicted.

s: the similarity score of chemical space for drugs or genomic space for targets

$N_d = \sum_{i=1}^{n_d} s_d(d', d_i)$

$N_t = \sum_{j=1}^{n_t} s_t(t', t_j)$

and its extension, NetLapRLS, where matrix $\mathbf{Y}$ is substituted with a network similarity matrix $\mathbf{X}$ whose (i,j)-th entry is the number of drugs interacting with target t_i and t_j or that of targets interacting with drug d_i and d_j to take interaction patterns into consideration so that reaction takes place for a drug-target pair (d_i, t_j) with high scores:

$$\bar{f}(d_i, t_j) = \frac{f(d_i, t_j) + f(t_j, d_i)}{2} \tag{2.16}$$

Comparing with BLM, NetLapRLS discover knowledge from both labelled and unlabelled samples but negative ones.

RLS-Kron classifier

Assuming that drugs with similar interaction profiles y_{d_i} on a set of targets tend to behave identically with an outside target, **?** construct the Gaussian kernel from the y_{d_i} for drugs:

$$K_{P,d} = exp(-\gamma_d ||y_{d'} - y_{d_i}||^2) \tag{2.17}$$

where the parameter γ_d can be determined with cross-validation to control the kernel bandwidth. Analogously for targets:

$$K_{P,t} = exp(-\gamma_d ||y_{t'} - y_{t_i}||^2) \tag{2.18}$$

Eqn(17, 18) are then combined with kernels representing chemical structure and genomic sequence information $K_{chem,d}, K_{bio,t}$ derived from similarity matrices using weighted average

$$K_d = \beta_d K_{chem,d} + (1 - \beta_d) K_{P,d} \tag{2.19}$$

$$K_t = \beta_t K_{bio,t} + (1 - \beta_t) K_{P,t} \tag{2.20}$$

and then integrated into an overall kernel for drug-target pairs using Kronecker product kernel [**????**] which allows predicting drug-target pairs concurrently with a lowered time complexity of $\mathcal{O}(n_d^3 + n_t^3)$.

However, the primary assumption limits its application on molecules or proteins with unknown interaction profile.

2.4 Matrix factorisation

A pioneering work to obtain a drug-target interaction inference scheme from a fully probabilistic perspective has been proposed by **?** using a Bayesian model that frames the interaction prediction problem as a matrix factorisation one. It combines binary classification and dimension reduction approaches by projecting drug molecules and target proteins into a unified low-dimensional subspace using kernels $\mathbf{K}_d, \mathbf{K}_t$ derived from chemical and biological data respectively and thus developing a score matrix for the final prediction matrix $\mathbf{Y}$. The rate-determining step is the calculations of covariance with an overall time complexity of $\mathcal{O}(RN_d^3 + RN_t^3 + R^3)$ in this variational inference mechanism.

Since Bayesian treatments are computational expensive, **?** further extend it utilising a probabilistic formulation to obtain the variational approximation efficiently and integrating multiple side information with different kernels at the mean time so that predictions on outside samples are allowed.

Chapter 3

Experiment

The kernel-based methods introduced above frame the prediction of drug-target interactions as three different problems:

1. Edge prediction

2. Binary classification

3. Matrix factorisation

Three models with distinct characteristics, RLS-Kron classifier[?], BLM[?], Bayesian model[?], have been selected to represent the three approaches respectively. In the following experiment, these three models will be implemented on the same dataset offered by ? to better compare performances of kernel based methods.

3.1   Data pre-processing

The datasets used in model evaluation is the Gold standard datasets released as the supplements of ISMB 2008 paper [?] whose interaction information covering four major drug-target interaction networks in human, enzymes, GPCRs, ion channels and nuclear receptors, is achieved from KEGG BRITE [?], BRENDA [?], DrugBank [?] and SuperTarget [?]. The numbers of chemical compounds, target proteins, their ratios as well as the known interactions are summarised below:

Subset	Drug	Target	Drug/Target	Positives	Positive/Negative
Enzyme	445	664	0.67	2926	
GPCRs	223	95	2.35	635	
Ion channels	210	204	1.03	1476	
Nuclear receptors	54	26	2.08	90	

Table 3.1: Summary of ISMB 2008 paper datasets

Similarities between each pair of small molecules d', d_i are formulated as the ratio between the number of nodes which are mutual substructures of both d' and d_i over the total number of nodes which are substructures of at least one of d' and d_i, i.e.

$$s_d(d', d_i) = \frac{d' \cap d_i}{d' \cup d_i} \tag{3.1}$$

Such measurement by the ratio of the size of intersection to the union of substructures sets is known as SIMCOMP[**?**].

In the case of proteins, normalised Smith-Waterman score[**?**] is adopted:

$$s_t(t', t_j) = \frac{w(t', t_j)}{\sqrt{w(t', t')}\sqrt{w(t_j, t_j)}} \tag{3.2}$$

As a result, two square matrices indicating similarity scores among drug compounds or target proteins will be the inputs of models.

3.2 Model evaluation

Parameters for each model are chosen from their original reports as concluded below:

RLS-Kron classifier	$\tilde{\gamma}_d = \tilde{\gamma}_t = 1, \alpha_d = \alpha_t = 0.5, \sigma = 1$
BLM	c=1
Bayesian model	$R = 25, \alpha_\lambda = \beta_\lambda = 1, \sigma_g = 0.1, \mu = 0$

Table 3.2: Parameter details

Cross validation techniques (CV) are considered to be statistically efficient in evaluating model performances. In K-fold CV, data will be split into k subsets in equal size. Each subset will be used as test data and the remaining data will be used as training data to train a model, that is, k models will be build. In this experiment, 10-fold CV which randomly split data into 10 subsets in equal size is employed in GPCR, ion channel and enzyme datasets while leave-one-out (LOO) CV, where k is always one less than the size of data, for Nuclear receptor dataset due to its small size. It is worth noticing that three models are trained on the same folds of chemical and genomic data as well as interaction information each time.

To quantify model performances, area under precision-recall curve (AUPR) is utilised rather than the most widely used measurement, area under receiver operating characteristic curve (AUC), on account of the large ratio of negatives to positives for interaction data. Besides each AUPR measure, p-values of paired t-test are marked to indicate how strong is the evidence for the contribution from the model achieving the best performance.

3.3   Results and discussion

Judged from AUPR values of drug and target prediction, it is easy to tell that LapRLS achieves the best performance, slightly outperforming Bayesian model and considerably for BLM. In terms of pair prediction, LapRLS still performed the best on enzyme and nuclear receptor datasets while it was not the case for GPCRs and ion channel where Bayesian model outperformed the other two. In general evaluation, target predictions are more accurate than that on drugs as a result of a more efficient similarity presentation for proteins whereas both of them are more difficult than pair prediction since new drugs or targets with unknown interactions are involved.

Space and time complexities quantify the execution time and memory space required by a program which are essential in efficacy measurement. Time complexity indicates the overall number of steps an algorithm takes to solve a problem whereas space complexity is measured by using polynomial amounts of memory with an infinite amount of time. Table 3.4 shows that kernel methods are computational

Targets	Dataset	RLS-Kron classifier	BLM	Bayesian mode
Enzyme	drug	0.368	0.092 (1.29×10^{-26})	0.253 ($3.09 \times 10^-$
	target	0.792	0.611 (7.92×10^{-25})	0.607 ($4.48 \times 10^-$
	interaction	0.826 (1.81×10^{-38})	0.496 (2.56×10^{-53})	0.796 ($4.12 \times 10^-$
GPCRs	drug	0.397 (5.01×10^{-1})	0.210 (6.50×10^{-29})	0.357 ($5.14 \times 10^-$
	target	0.508	0.367 (2.90×10^{-15})	0.412 ($1.01 \times 10^-$
	interaction	0.640 (3.43×10^{-29})	0.464 (4.03×10^{-35})	0.686 ($1.03 \times 10^-$
Ion channel	drug	0.366	0.167 (1.51×10^{-20})	0.296 ($2.15 \times 10^-$
	target	0.778	0.641 (2.26×10^{-21})	0.725 ($6.90 \times 10^-$
	interaction	0.804 (1.08×10^{-38})	0.592 (1.45×10^{-39})	0.876 ($2.72 \times 10^-$
Nuclear receptor	drug	0.482 (1.79×10^{-2})	0.194 (3.49×10^{-20})	0.450 ($3.46 \times 10^-$
	target	0.432 (4.47×10^{-1})	0.325 (3.10×10^{-3})	0.404 ($2.19 \times 10^-$
	interaction	0.539 (3.29×10^{-5})	0.204 (2.09×10^{-24})	0.508 ($1.37 \times 10^-$

Table 3.3: AUPR values by 5 trial 10-fold CV with p-values

Model	Space	Time	
		Train	Test
RLS-Kron classifier	$\mathcal{O}(n_d n_t)$	$\mathcal{O}(n_d^3 + n_t^3)$	$\mathcal{O}(1)$
BLM	$\mathcal{O}(n_d n_t)$	$\mathcal{O}(n_d n_t (n_d^2 n_t^2))$	$\mathcal{O}(1)$
Bayesian model	$\mathcal{O}(n_d n_t)$	$\mathcal{O}(R(n_d^3 + n_t^3)t)$	$\mathcal{O}(1)$

Table 3.4: Complexities comparison

expensive which significantly restricting their utility on large datasets.

Chapter 4

Improvement

To circumventing the local minimum issue, the application of kernel methods is generally formulated as a quadratic programming (QP) problem whose time and space complexities of $\mathcal{O}(N^3)$ and $\mathcal{O}(N^2)$ respectively when implemented in a straightforward way are computationally infeasible. Much effort has been devoted to fast learning approaches for large data sets from various perspectives. In this section, two developmental milestones, generalised core vector machine (CVM) and fast kernel density estimate (FastKDE), are reviewed succinctly.

4.1 CVM

The CVM algorithm[?] simplifies kernel methods into equivalent minimum enclosing ball (MEB) problems so that not only time complexity is linear in N and the space complexity independent of N but also approximations converge to optimality. A generalised CVM [?] in the form of

$$
\begin{aligned}
\text{maximise} \quad & \boldsymbol{\alpha}^T(diag(\boldsymbol{K}) + \Delta) - \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha} \\
\text{s.t.} \quad & \boldsymbol{\alpha}^T \mathbf{1} = 1, \alpha \geq 0
\end{aligned}
\tag{4.1}
$$

extends the applicability for a larger range of kernel methods.

4.2   FastKDE

Unlike CVM, FastKDE innovatively scales up kernel methods in a perspective of kernel density estimation (KDE) bridged by an entropy-based integrated squared error ISE criterion which expresses the problem as finding a set of variable weights $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_N\}$, the minimiser of a weighted-sum objective:

$$
\begin{aligned}
\text{minimise} \quad & J_1 + \lambda J_2 = ISE(\alpha) + \lambda \hat{V}(\boldsymbol{S}; \boldsymbol{\alpha}) \\
\text{s.t.} \quad & \boldsymbol{\alpha}'\boldsymbol{1} = 1, \boldsymbol{\alpha} \geq 0
\end{aligned}
\tag{4.2}
$$

where:

$\hat{V}(\boldsymbol{S}; \boldsymbol{\alpha})$: a quadratic entropy estimator of the probability density function

λ: a parameter gives relative weight between J_1 and J_2

As proved by **?**, KDE with any kernel variants can be derived such that the optimisation problem can be further formulated as

$$
\boldsymbol{\alpha} = \arg \min \boldsymbol{\alpha}'\boldsymbol{K}\boldsymbol{\alpha} - \alpha' \text{diag}(\boldsymbol{K} + \Delta)
\tag{4.3}
$$

where:

$\Delta = \frac{(1+\lambda)\text{diag}(\mathbf{K})}{1-\lambda}$

A sampling strategy is then adopted to find a new KDE on a reduced data set with the estimation accuracy suffering little:

$$
\begin{aligned}
\text{minimise} \quad & D(\hat{p}(x), \tilde{p}(x)) \\
\text{s.t.} \quad & \tilde{p}(x) = \frac{1}{m} \sum_{i=1}^{m} k(x, x_i)
\end{aligned}
\tag{4.4}
$$

Chapter 5

Conclusion

This article reviews state-of-art kernel based methods for predicting drug-target interaction. Three typical methods framing the prediction task into different problems have been implemented on the same gold standard dataset, the experimental results of which shows an overall outperforming of LapRLS, followed by Bayesian model and further by BLM. The unknown interactions of new compounds or proteins complicate predictions on drug or target data so that AUPR values for pair predictions always achieves the highest. The universal lower AUPR values on drug predictions than that of targets might be a result of a more efficient similarity presentation of proteins. More efforts will be devoted from three perspective:

Bibliography

Basilico, J. and Hofmann, T. (2004). Unifying collaborative and content-based filtering. *ICML '04 Proceedings of the twenty-first international conference on Machine learning*, page 9.

Ben-Hur, A. and Noble, W. S. (2005). Kernel methods for predicting protein–protein interactions. *Bioinformatics*.

Bleakley, K., Biau, G., and Vert, J. P. (2007). Supervised reconstruction of biological networks with local models. *Bioinformatics*, 23(13):i57–i65.

Bleakley, K. and Yamanishi, Y. (2009). Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics*, 25(18):2397–2403.

Erhan, D., L'Heureux, P.-J., Yue, S. Y., and Bengio, Y. (2006). Collaborative filtering on a family of biological targets. *Journal of chemical information and modeling*.

Faulon, J. L., Misra, M., Martin, S., Sale, K., and Sapra, R. (2008). Genome scale enzyme–metabolite and drug–target interaction predictions using the signature molecular descriptor. *Bioinformatics*.

Gönen, M. (2012). Predicting drug–target interactions from chemical and genomic kernels using bayesian matrix factorization. *Bioinformatics*, 28 (18).

Gonen, M., Kaski, S., and Member, S. (2014). Kernelized bayesian matrix factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10):2047–2060.

Günther, S., Kuhn, M., Dunkel, M., Campillos, M., Senger, C., Petsalaki, E., Ahmed, J., Urdiales, E. G., Gewiess, A., Jensen, L. J., Schneider, R., Skoblo,

R., Russell, R. B., Bourne, P. E., Bork, P., and Preissner, R. (2008). Supertarget and matador: resources for exploring drug-target relationships. *Nucleic Acids Research*, 36(919-922).

Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M. (2003). Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *Journal of the American Chemical Society*, 125(39).

Hopkins, A. L. and Groom, C. R. (2002). The druggable genome. *Nature Reviews Drug Discovery*.

Hue, M. and Vert, J.-P., editors (2010). *On learning with kernels for unordered pairs*. ICML.

Jacob, L. and Vert, J. P. (2008). Protein-ligand interaction prediction: an improved chemogenomics approach. *Bioinformatics*.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in kegg. *Nucleic Acids Research*, 34(354-357).

Mordelet, F. and Vert, J. P. (2008). supervised inference of regulatory networks. *Bioinformatics*, 24(16):i76–i82.

Oyama, S. and Manning, C. D., editors (2004). *Using Feature Conjunctions across Examples for Learning Pairwise Classifiers*. Proceedings of the 15th European Conference on Machine Learning (ECML2004).

Schölkopf, B., Tsuda, K., and Vert, J.-P. (2004). *Kernel methods in computational biology*. Cambridge, Mass. : MIT Press.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. (2004). Brenda, the enzyme database: updates and major new developments. *Nucleic acids research*, 32.

Schöölkopf, B., Smola, A., and Müüller, K.-r. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*.

Tsang, I., Kwok, J. T. Y., and Cheung, P. M. (2005). Core vector machines: Fast svm training on very large data sets. *Journal of machine learning research*, 6:363–392.

Tsang, I. W. H., Kwok, J. T. Y., and Zurada, J. M. (2006). Generalized core vector machines. *IEEE transactions on neural networks*, 17(5):1126–1140.

Van Laarhoven, T., Nabuurs, S. B., and Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics*.

Wang, S., Wang, J., and Chung, F.-L. (2014). Kernel density estimation, kernel methods, and fast learning in large data sets. *IEEE Transactions on Cybernetics*, 44(1):1–20.

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008). Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, 36.

Xia, Z., Wu, L.-Y., Zhou, X., and Wong, S. T. (2010). Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Systems Biology*.

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*.

Zhu, S., Okuno, Y., Tsujimoto, G., and Mamitsuka, H. (2005). A probabilistic model for mining implicit 'chemical compound-gene' relations from literature. *Bioinformatics*.